

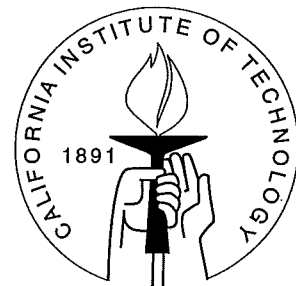
DIVISION OF THE HUMANITIES AND SOCIAL SCIENCES

CALIFORNIA INSTITUTE OF TECHNOLOGY

PASADENA, CALIFORNIA 91125

RULES FOR EXPERIMENTING IN PSYCHOLOGY AND ECONOMICS, AND WHY THEY DIFFER

Colin Camerer



SOCIAL SCIENCE WORKING PAPER 946

December 1995

Rules for Experimenting in Psychology and Economics, and Why They Differ

Colin Camerer

Abstract

This chapter discusses methodological differences in the way economists and psychologists typically conduct experiments. The main argument is that methodological differences spring from basic differences in the way knowledge is created and cumulated in the two fields – especially the important role of simple, formal theory in economics which is largely absent in psychology.

JEL classification numbers: 036, 215

Key words: Experimental economics, psychology, bounded rationality.

Rules for Experimenting in Psychology and Economics, and Why They Differ

Colin Camerer¹

1 Introduction

There are many ways to conduct experiments in social sciences. This chapter contrasts two philosophies or approaches, referred to as the "experimental economics" (E) and "experimental psychology" (P) approaches. The distinction is not sharp, of course, but even fuzzy distinctions can be helpful in framing debates and clarifying misunderstanding. The main focus in this chapter is the difference between E and P experiments in research in decision theory and game theory. There is substantial bickering between these camps about the right way to conduct experiments. My central argument in this chapter is that many crucial differences in experimental style are created by underlying differences in working assumptions about human behavior and about how knowledge is best expressed, not by underlying differences in areas of substantial interest. My hope is that understanding this point can sensitize both sides (and outsiders or newcomers) to differences in these philosophies, without particularly advocating one over the other, in order to defuse some sensitive aspects of the debate between E and P.

The chapter has several sections. Next I argue that economics and psychology are more sharply distinguished by style than by substance. In section 2 we make the leap to experimental style, describing key differences in P and E experimental styles and showing how they follow from differing presumptions in the underlying disciplines. Section 3 gives a specific example to further illustrate the argument. In section 4 I express some preferences for certain stylistic features, add caveats, and mention the exceptional methodological style of Reinhard Selten.

This paper draws clearly on many aspects of Smith's classic 1976 and 1982 papers which guided experimental work in economics for the last twenty years. I was also inspired by a thoughtful unpublished 1988 paper by Daniel Kahneman which contrasted P and E methods. Similar ground is covered by Cox & Isaac (1986). Descriptions of now-standard methodology in experimental economics can be found in Hey (1991), Davis & Holt (1992), and Friedman & Sunder (1994). Chapters in Kagel & Roth (1995) review discoveries in experimental economics.

¹ Comments from participants at the Big Ten Accounting Doctoral Consortium, Minneapolis 5/13-15/94, Chip Heath, Wulf Albers, and Werner Guth, were helpful. Many conversations over the years-- particularly with Daniel Kahneman, Charlie Plott, and Richard Thaler, and numerous coauthors-- have shaped the contents of this paper. Discussions at the Russell Sage Foundation during my 1991-92 visit there were especially formative. Forthcoming in W. Guth and E. Van Damme (eds.), *Essays in Honor of Reinhard Selten*. Berlin: Springer-Verlag.

1.1 Substance vs. Style: Economics and Psychology

I claim differences in E and P experiments arise from differences in style of the two fields. By substance I mean, "what is the basic question being asked?" By style, I mean, "What constitutes an answer to the question? How are answers expressed?"

As a starting point for discussion, let's suppose psychologists ask:

P: How do people think & behave?

and economists ask

E: How are scarce resources allocated?

At first blush, this difference in fundamental "why" questions seems to sharply divide the two fields along substantive lines-- e.g., psychologists study human thinking, and economists study resource allocation.

But the two questions also express inherent methodological differences. Answering the P question demands observation of individual subjects in controlled settings (experiments). The E's "allocation of scarce resources" hints strongly at an underpinning from mathematical optimization. My claim is that the substantive difference is actually small, and the stylistic difference large.

1.2 How much substantial overlap is there?

Take any one of several concrete examples: How do people decide how much time to spend with their families, instead of working? Why do people discriminate against others who are unlike them? Who do people vote for? When do people retire? How do people choose friends? Spouses? Why do people break the law? How do people invest their wealth? Choose jobs or education? Why do people become addicted to drugs and acquired tastes?

Virtually all these questions can be posed as P or E questions, and have been studied by social scientists of both types (and also by sociologists, whose work is ignored in this chapter for brevity). To further illustrate that the differences between P and E are not entirely substantial, Table 1 illustrates some of the similarities in topics of substantial interest across the two fields. Many interesting topics-- altruism, learning, taste formation-- have been studied by people in both fields. The key difference is perhaps not the substantive topics psychologists and economist study, but the

way theories are expressed and data are gathered.

Table 1: Similarities in substantial topics in P and E

<u>Psychology</u>	<u>Economics</u>
<u>GUIDING QUESTIONS</u>	
How do people think & behave. Why?	How are scarce resources allocated?
<u>SPECIFIC TOPICS</u>	
reciprocity norms	"gift exchange" labor models
prisoner's dilemma	externalities
expectancy theory	expected utility (EU & SEU)
altruism	altruism (bequests to children)
impulsivity	impatience (discounting)
language & meaning	language & meaning (cheaptalk)
learning	learning (least-squares, population learning)
taste formation	taste modelling (addiction, participation externalities)
symbolic consumption	signaling
conformity/herd behavior	info. cascades, rational conformity

1.3 Stylistic differences in E and P

I think the fundamental stylistic differences are two: (i) E's strongly prefer mathematical formalisms. And since formalisms become unwieldy and inelegant as they get too complex, E's like simple formal models. (ii) E's like predictions that are surprising, and testable using naturally-occurring ("field") data. Note how features (i) and (ii) work together. The desire for parsimony enables many E's to justifiably ignore whether assumptions are realistic (if alternative assumptions are "too" complicated), and hence place more value on naturally-occurring data than on experimental refutations.

In contrast, P's tend to prefer verbal models or principles and eschew sophisticated mathematical expressions. Since these principles need not fit together in mathematically coherent ways, models can resemble lists of effects or critical variables, and can be expressed (and discovered) by experimental demonstrations. Field tests or demonstrations are consequently rare.

My argument for stylistic difference implies that a single substantive question can be translated from the P to E mode, or vice versa.

For example, translating from P to E, it is a simple matter to translate a loose psychological

account of, say, how financial incentives influence performance into economic terms: The "scarce resource" is attention or thinking effort, higher incentives produce more thought (the supply curve for thinking is upward-sloping) and more thought improves performance (thought is a factor or production in generating performance), as in Smith and Walker (1993). But doing so has no special interest to many Ps, because the resulting model will of course be oversimplified ("reductionist") and plainly at odds with at least some experimental observations.

Similarly, one can ask an E question--do people play the subgame perfect solution to a sequential bargaining problem?-- in a P way, by observing whether subjects in a laboratory search for information in the way that is necessary to calculate perfect equilibrium using backward induction (Camerer, Johnson, Sen & Rymon, 1993). The answer to that question is "No", but that answer does not deter an E from continuing to assume the answer is yes, if mathematical simplicity is sufficiently important and no equally parsimonious replacement assumption emerges from the experimental findings.

A helpful and charitable way to think of the E and P camps is as competing cultures, religions, or schools of artistic expression. Psychologists are realists who paint detailed landscapes and strive to make them "lifelike". (A good psychology experiment is a drama that makes a point about human nature.) Economists are abstract expressionists who value pictures that expresses a feature of the world with minimal clutter. Naturally, the P's criticize the E's for clinging to unrealistic assumptions, and the E's criticize the P's for lacking formal theory, or parsimony.

2 Experimental styles in E and P

My central claim is that general stylistic differences in how knowledge is generated in the E and P fields accounts for many of the major differences in how experiments are conducted.

Table 2 lists some important differences in experimental styles. Most of the differences spring from basic differences in media for expression of knowledge and how the interplay of theory and data work in the two fields.

For E's, the medium for expression of knowledge is, primarily, a body of formalism and modelling principles. A body of facts buttresses the formalism and sustains faith in it, but is clearly subservient because ample evidence against a formal principle is often insufficient to create widespread rejection and search for replacements. (E.g., the experimental evidence against expected-utility maximization is now forty years old and still piling up, but textbooks and most applied work continue to assume it.) The interplay of theory and data occurs as good experiments test theory or provoke its development by exploring behavior in an interesting domain where theoretical principles haven't been developed. For P's, theory is sheer cloth wrapped around a body of prominent

experimentally-observed results. Explanations that don't fit new observations are jettisoned more quickly than old facts are.

Several properties of good experimental design follow from these basic differences.

Context or labels: In experimental E, an abstract context (or bland verbal labels) is sufficient to test theory; it is considered unnecessary to make the experimental context realistic since most theory does not predict that results will depend on verbal labels. (Indeed, labelling choices may even be undesirable since it may interfere with attempts to clearly control preferences by paying financial incentives.) By contrast, in experimental P a rich context is often used to motivate subjects to behave realistically and to provide a concrete illustration of an underlying (mathematical) principle. In addition, most psychologists feel that abstract, symbolic information is processed differently than concrete, lifelike stimuli. Hence, learning about subjects' intuitive grasp of statistics using only dice and coins could be misleading as a guide to how subjects reason about baseball statistics or categorize people into stereotypes.

Subject pools: Economic theories usually do not predict any reason why different pools of subjects will behave differently, so E's rarely collect data on the demographics of subjects or test whether behavior varies with gender, age, education, etc. (though see Ball & Cech, 1990, for a review). Indeed, nonhuman subjects such as rats, pigeons, et al have been productively used in experiments on consumer choice theory, theories of labor supply, and expected utility (see Kagel, Battalio & Green, 1995). A long (now quiet?) tradition of interest in personal differences by some psychologists leads most psychologists to at least record personal information, if not theorize about differences due to age, gender, educational background, etc.

Types of data: Since economic theory usually does not predict what cognitive process will lead to an outcome (say, competitive equilibrium in a market or a game-theoretic equilibrium), E's usually do not collect types of data other than choices. In contrast, data like Likert rating scales of attractiveness, response times, self-reports by subjects of what they were thinking, post-experiment justifications for their choices, etc. are widely used in psychology and theories often make predictions about these variables. (Theory-minded E's would not know what to do with such data; though some E's interested in decision error have begun to use response times.)

Repetition of trials: Economic theories are usually asserted to apply to equilibrium behavior, after initial "confusion" disappears and (unspecified) equilibrating forces have had time to operate.

Table 2: Differences in P and E Experimental Style

<u>Psychology (P)</u>	<u>Economics (E)</u>
MEDIUM FOR EXPRESSION OF KNOWLEDGE	
Body of facts (1st) Informal interpretation	Formal model (1st) Body of facts
INTERPLAY OF THEORY & DATA	
Experimental results are cloth from which theory is woven	Experiments operationalize and test general theory in specific (artificial) setting
CONTEXT/LABELS	
Rich context preferred: engages subjects more lifelike	Rich context avoided: creates nonpecuniary utility irrelevant from theory's view
SUBJECT POOLS	
Gender, age, etc. often recorded	Theory predicts no effects (or abstractness of setting swamps subject pool differences)
TYPES OF DATA	
Ratings, response times, subject self-reports, etc.	Choices. Self-reports not informative.
REPEATED TRIALS?	
No need. Some of life is like first trial	Yes. Theory predicts "last" (equilibrium) trial
INCENTIVES OF SUBJECTS	
Volunteer subjects are well-motivated to "try hard"	Financial motivation best.
DATA ANALYSIS & REPORTING	
Assumes friendly reader. Reduced-form. Sometimes F's or p-values only. Don't report "failures".	Assumes skeptical reader. More raw data. (Reader can reanalyze). No "failures" in a well-designed (competing-theory) experiment.
DECEPTION?	
Sometimes, to create unnatural situations or break confounds.	Rarely. (Experimenter credibility is a public good)

So E's generally conduct an experiment for several repeated trials under "stationary replication"--destroying old goods and refreshing endowments, in order to recreate the decision situation in exactly the same way each time (except for anything that was learned from the previous experience, in a "Groundhog Day" design²). Then special attention is paid to the last periods of the experiment (e.g., often the last half of the trials are used as data to test an hypothesis) or to the change in behavior across trials. Rarely is rejection of a theory using first-round data given much significance. P's, in contrast, sometimes run a single trial for each stimulus if learning is of no special interest.

Incentives of subjects: A crucial property of most E experiments is that the consequences of their choices are made clear (or "salient") to subjects, and they are always given financial incentives. The standard explanation for insisting on salience of this sort, and financial rewards (rather than say points, course credit, hypothetical dollars, etc.) is articulated by Smith (1976, p. 277):

"Individuals may attach game value to experimental outcomes...Because of such game utilities it is often possible in simple-task experiments to get satisfactory results without monetary rewards by using instructions to induce value by role-playing behavior...But such game values are likely to be weak, erratic, and easily dominated by transactions costs, and subjects may be readily satiated with 'point' profits."

The insistence on paying subjects is absolute among experimental E's even though most, in their hearts and in hallway discussions, are agnostic about whether results would differ substantially if subjects' payments were hypothetical.³ The insistence on money seems to be a fetish arising from the frequently-made assumption that people are self-interested and largely motivated by money. My view is that more comparisons of high- and low-incentive conditions (including hypothetical payment) are needed to test Smith's assertions. Probably the result of such studies would be that in some situations, where the task is either particularly hard or simple (or intrinsic motivation is high), paying performance-based incentives makes little difference. In tasks which get boring quickly (e.g., "probability matching"), or where performance is of intermediate responsiveness to effort, paying subjects could make a lot of difference. Also, many studies comparing hypothetical and actual payments have found that the frequency of outliers is reduced by paying real money, which may be particularly important in outlier-sensitive situations, like minimum-action coordination games, or

² In the movie "Groundhog Day" a hapless reporter is forced to repeat the same 24-hour sequence of interactions over and over, with everything unchanged except his own memory of the outcomes of his actions in the previous day. The endless looping of experience, while boring, turns out to be ideal for learning since different actions can be tried while everything else is held equal. This is precisely the "stationary replication" design widely used in experimental economics.

³ It is virtually impossible (I know of no examples in the last ten years or so) to publish experimental results in a leading economics journal without paying subjects according to their performance. Paying subjects often costs substantial sums of money, which acts like a very high submission fee, and may particularly limit the ability of younger investigators or those in less well-funded institutions to generate widely-read results. Also, paying subjects or using incentive-compatible mechanisms (like the "lottery ticket" procedure for inducing risk tastes, or the Becker-DeGroot-Marschak procedure for selling prices) can complicate an experiment or produce demand effects.

markets for long-lived assets which can exhibit price bubbles.

In contrast, experimental P's usually do not pay subjects (though some do); instead they rely on the intrinsic motivation of subjects who volunteered out of curiosity, and whose attention is sustained by pride and eagerness to do a good job. (Unfortunately, in my view, many subjects in introductory P classes are forced to participate in experiments as part of course credit; they may indeed be readily satiated or only erratically motivated by points, as Smith feared.)

Data reporting and analysis: Styles of data analysis are different in E and P. Experimental E's generally follow the useful convention in reporting other kinds of empirical economics (e.g., econometric results), report lots of raw data, using clever graphical methods and often testing hypotheses in statistically sophisticated ways (see, e.g., El-Gamal & Palfrey, in press, on Bayesian methods and "optimal design"). Then readers, knowing as much about the data as space permits, can form their own opinions about the results rather than blindly accepting the author's.

In my opinion, data reporting and analysis standards are simply lower in many domains of experimental P (particularly social psychology). P's often report highly aggregated data, using embarrassingly simple bar charts, with outliers trimmed capriciously and analyzed with crude statistical tests that are usually parametric and often badly misspecified. (F-tests are routinely used for differences in variances of grouped data that are clearly non-normal; tests assuming continuous variables are often applied to discrete data like 5- or 7-point scale ratings.) There also appears to be more of a "file drawer" problem in psychology-- only surprisingly large effects are reported, ensuring that exact replications will show smaller effects due to regression-toward-the-mean. (Some psychologists publish only a third or fourth of the data they collect, which suggests the magnitude of the file drawer problem that can result if findings are not carefully replicated before publication.)

While my description of experimental P data analysis methods is harsh, I think they can again be understood as partly forgivable in the light of other features of psychological research style. Experimental P's often have access to large subject pools at low cost (since subjects are often not paid at all), so free subject labor can substitute for expensive experimenter human capital in designing efficient statistical tests. And in experimental P, the experimenter writes a script (perhaps a short vignette, like a joke) or lets a drama unfold that illustrates an informal claim about human nature. Readers give the experimenter the benefit of the doubt (trusting her results and not needing to see raw data), much as theatergoers trust the playwright. In this view, the file drawer effect results from refusing to stage bad plays.

Deception: The use of deception is another important stylistic difference between E's and P's: E's hate it, and P's often don't mind. For the sake of discussion, define deception as actively claiming something to be true, when you know it to be false. For example, some experimenters have implicitly promised to pay subjects money in a way that (saliently) depended on choices, then told them at the end they would earn a flat amount. In other cases a confederate subject, playing a role necessary to

create a social illusion or treatment, is implicitly introduced as a fellow subject, experimenter, former subject, etc.

Experimental E's hate deception because they feel it is often unnecessary, and more importantly, harms the credibility of all experimenters. Since American Psychological Association rules require experimenters to inform a subject about any deception in a post-experiment debriefing, veteran subjects always know they have been lied to before. Some studies also indicate that subjects who are told about a deception in a debriefing do not always believe the debriefing either. (Should they?) An example will help illustrate some features of this disagreement about deception.

Many experimenters have studied "ultimatum" games in which one person offers a division of a sum X to a responder. Then the responder either takes it, or rejects it and leaves both with nothing (see Guth et al, 1982; Camerer & Thaler, 1995). Suppose you were interested in whether a sense of "entitlement" affected the kinds of divisions players offered and accepted. Hoffman et al (1994) studied this by having pairs of subjects play a simple game of skill. Those who won got to make the offer in a subsequent ultimatum game; those who lost could accept or reject the winner's offer. They found that winners offered less (keeping more for themselves) and losers were willing to accept less. Winning created entitlement.

Or did it? There is a subtle problem with this design: It perfectly correlates game-playing ability and entitlement generated by winning. Their results are consistent with the hypothesis that skilled game-players always demand more, unskilled players always accept less, and game-playing simply sorted players by their bargaining aggressiveness (or timidity) rather than created entitlement per se. For a psychologist, the natural way to break this two-variable confound is to play a game, then randomize feedback on whether players won or lost. That way, half the skilled game-players land in the low-entitlement group (they are told, deceptively, they lost); and half the unskilled players land in the high-entitlement group. Deception is required in this alternative design; it is hard to see how to break the skill-entitlement confound otherwise. One can see how experimental P's might criticize the Hoffman et al design for allowing a confound, and if one is eager enough to truly separate skill and entitlement, deception may be scientifically useful.

This example illustrates the scientific benefit of deception. But it is important to take account of the possible costs of deception and weigh costs and benefits. Two costs loom largest: First, repeated deception probably harms overall experimenter credibility, but the extent of harm and whose experiments are jeopardized is not well understood. Second, since deception is generally used to create an unnatural social situation-- such as a non-relation between skill and entitlement-- one must ask: If the situation being created is sufficiently unnatural to require deception, will subjects believe the deception? If subjects entertain the hypothesis that they are being deceived, then the most necessary deceptions-- to produce the most unnatural situations-- are least likely to be believed. This point can be illustrated by further ultimatum experiments.

Polzer, Neale & Glenn (1993) (PNG) used an ultimatum game design that did break the confound between skill and entitlement. After their subjects did a "word find" task, they were given deceptive feedback about actual performance, in order to sort half the high-skilled players into a low-entitlement condition (by telling them they performed more poorly than the subject they were paired with). PNG reported that entitled ("justified") players offered a modest \$.32 less than nonentitled players (out of \$10 being divided). More interestingly, the responder subjects who were nonentitled (were told they had low performance) demanded \$.41 more than entitled players did. A natural possibility is that the deceptive feedback was not believed by some of the players who were told they were less skilled. Evidence for this possibility is the fact that the surprising increase in demands of nonentitled players is larger among groups of (self-selected) friends, who came to the experiment together and, knowing more about each other, were more likely to doubt the false feedback, than among groups of strangers.

3 An example: Fairness in surveys and markets

An example will help illustrate some of the differences in experimental E and P styles summarized in Table 2.

3.1 Surveys of fairness

Many social scientists have been interested in perceptions of fairness and their impact on social life. Kahneman, Knetsch & Thaler (KKT) (1986) set out to describe some of the rules people use to judge whether economic transactions are fair. They decided to use surveys in which short vignettes were described, and respondents could answer whether a transaction was fair or not. For example, subjects were told:

A hardware store has been selling snow shovels for \$15. The morning after a large snowstorm, the store raises the price to \$20. Please rate this action as:

Completely Fair Acceptable Unfair Very Unfair

Of 107 respondents, 82% said the store's price increase was unfair or very unfair. From dozens of questions like this, they induced several basic propositions about perceptions of fairness.

Their method illustrates some classic features of experimental P. The survey data are used to construct theory, not test it. The natural context makes the question more engaging than a drier,

abstract version ("Is it fair to mark up prices after shortages?") would be, and perhaps easier to comprehend as well. The subjects were randomly chosen and telephoned, because KKT thought that average folks were obviously the people whose behavior they wanted to model, and students might be unfamiliar with many of the contexts they asked questions about (e.g., employers raising or cutting wages). The data are answers to questions, rather than choices with financial consequences. No repetition is necessary (though many different questions were asked) since nothing much can be learned from asking the same question twice. The data analysis is simple (and in this case, need not be complicated). No deception is involved.

From an experimental E view, these data tell us nothing about "real" behavior with substantial consequences for respondents. But the data are clearly useful for theory construction if one presumes only that (i) subjects are intrinsically motivated to answer accurately (i.e., to report what they really would feel about fairness if they rushed in from the snow and picked up a shovel marked \$15), or have no reason not to; and (ii) their perceptions of fairness have something to do with their willingness to forego buying a shovel they think is priced unfairly. A critic who thinks the answers are worthless is implicitly assuming that either (i) or (ii) are false, which reflects a judgment about human behavior that itself seems empirically false to psychologists who have experience with survey design.

3.2 Fairness in markets

Inspired by the KKT results, Kachelmeier & Limberg & Schadeewald (KLS 1991) designed a traditional E experiment to see whether fairness effects would affect prices and quantities in a market. Their clever design starts with the proposition, derived by KKT from their surveys, that a price increase is fair if it is accompanied by an increase in marginal costs, and is unfair otherwise. (For example, raising the snowshovel price to \$15 after the snowstorm-induced demand shock is unfair-- according to respondents to the question above-- but subjects also said raising the retail price of lettuce is fair after a transportation mixup created a shortage and raised the wholesale price.)

Their design began with several periods of posted-bid trading in which each of several buyers posted a bid at which they would buy a unit of an unspecified commodity from a seller. (This unusual design enabled direct observation of how buyers reacted to a visible cost change, the closest analogy to the KKT surveys.) Then partway through, a tax on seller revenues was imposed. The tax effectively shifts the marginal cost curve (since the price at which a seller must sell to earn zero after-tax profit is now higher).

In a control condition, the buyers knew nothing about the tax; the fairness theory predicts they may object to price increases in this case. In the cost-disclosure condition, buyers were told about

the revenue tax and it was explained why this might raise prices. Fairness theory predicts they will treat price increase as more fair in this case, and prices will rise more rapidly. In a third profit-disclosure condition, buyers were told the effect of the tax on the sellers' share of total surplus (at various possible prices). By design, the revenue tax also raised the sellers' share of total surplus at every price, so fairness theory predicts that buyers might resist paying higher prices in this case.

As the survey results implied, behavior in the three disclosure conditions was different. When the cost effect was disclosed, prices rose more rapidly to the new equilibrium price (buyers appeared to forgive price increases, or treated them as more fair). When the unfair profit effect was disclosed as well, price increases were slower. The no-disclosure control condition prices fell in between. Deng et al (1994) replicated the KLS findings using a more traditional posted-offer institution, and found similar convergence patterns but little difference in equilibrium prices across the information treatments.

The ingenious KLS design takes the KKT results seriously but plays by experimental E rules. The experiment tests theory, in an abstract context (making the influence of fairness all the more remarkable and plain-- it stems purely from attitudes about divisions of money, not about shovels or lettuce). Subjects make choices (bids and sales) with financial consequences, over groundhog-day repeated trials to see if fairness effects "wear off" (they do, to some extent). The time series of data are reported, so readers can judge how quickly equilibration occurs, and tests for differences in conditions are sophisticated (and conservative). No deception is involved.

The KKT-KLS interplay also shows how much can be learned by cross-fertilization of the E and P approaches. The KLS design would never have come about if many P-style survey questions did not reveal the texture of perceptions of fairness, and provide a clear hypothesis testable in later work. At the same time, survey questions are not well-designed, as the experimental market is, to test how much people will pay to punish unfair behavior, and how their behavior changes with repetition. So the two methods are productive complements in producing a fuller understanding of fairness in economics.

4. Conclusions and caveats

4.1 Which approach is better?

My own tastes find something of special value in each of the E and P approaches that could be usefully imported into the other.

The modelling tradition in E, and guidance from theory, imposes useful discipline on

economics experiments. Designing an experiment with a theory in mind forces you to be crystal clear about why all the pieces of the design are there. (It also clarifies theory, by forcing a theorist to be clear about at least one concrete domain in which the theory applies.) Designing to test theory has enabled us to create extraordinarily efficient experiments in domains where multiple theories compete. For example, Schotter, Weigelt & Wilson (1994) and Banks, Camerer, Porter (1994) each constructed a series of games to test several different equilibrium concepts. Designs of such efficiency are rare in psychology, at least partly because specific theories which predict point estimates or different directional responses to variables are less common.

Experimental E's are also extremely good about reporting raw data (many articles include raw data and instructions, to make reanalysis or replication by readers easy).

On the other hand, many precepts in the E approach are restrictive. Ignoring almost all data but choices wastes a valuable opportunity to learn something more from a group of subjects who are often eager to explain their thinking processes and inferences. (Whether their thoughts are useful or not is difficult to answer, but it is surely less difficult to answer if we collect such data!) Also, many types of apparently non-choice behavior are choices and should be given equal or comparable status. For example, the time it takes a subject to respond in making a choice is itself a choice (of how much time to spend answering the question); theories of decision cost might make predictions about this time-choice.⁴

Insisting on paying subjects according to performance may miss the opportunity to learn something by varying incentives (see Smith & Walker, 1993), impose special burdens on poorly-funded investigators, and complicate some experiments unnecessarily.

In addition, the distaste for "rich", naturally-labelled contexts has inhibited the scope of problems people have studied (particularly those lying most squarely at the crossroads of psychology and economics); though happily, I think this stricture has weakened substantially in recent years.

4.2 Some caveats

I end with some caveats. Obviously, there are many experimental E's and P's who do not fit well into the Table 2 columns, or fit in the "wrong" one. Mathematical psychologists are few in number but respect formalism and contribute substantially to it. Cognitive psychologists often study abstract tasks and offer modest incentives (perhaps to alleviate boredom). Some psychologists, like Amnon Rapoport, Ward Edwards, and Robyn Dawes, play by experimental economics rules for the same reasons E's do-- to force subjects to take tasks seriously, to make incentives (and hence

⁴ There is a substantial literature in psychology on response times (for example, they are often used as a primary dependent variable in theories of memory). Among decision theorists, John Hey, Nat Wilcox, and others have found them useful to study.

performance) salient, to study the effect of learning over repeated trials, etc.

And of course, some groups of experimental E's are more curious about psychological regularities, and less inclined to express findings in narrow formalisms. Prominent among such eclectic E's are members of the "German school".

Which brings me to Reinhard Selten, who founded and inspired the German school along with Heinz Sauermann. I wrote this essay for a book honoring the remarkable Reinhard Selten because his own work combines the very best of both the E and P approaches. But how?

Selten does it by having an open mind, reading and thinking voraciously, and by taking a "dualist" position. The dualist position is that formal theories of rational behavior are important for answering sharply-posed analytical questions, and possibly as normative bases for ideal living, but it is silly to think such theories are the best possible account of complicated behavior of actual humans. Instead, rational theories and behavioral (descriptive) theories must flow along separately.

Consequently, Selten and others in the German school mix and match conventions of the experimental E and P approaches as they see fit. Selten's experimental papers reflect the P's characteristic attention to details of subjects' decision rules and individual differences. For example, Selten developed the "strategy method", a survey/choice hybrid in which subjects report what they would choose (and then are forced to) for each of several possible realizations of a random variable, like a private value in an auction, or a randomly-determined Harsanyi "type" in an incomplete-information game. For most purposes, the strategy method produces much richer data and enables within-subject tests which are much more statistically efficient than single-choice methods.

To achieve mastery of a single mathematical approach or experimental style in a lifetime is difficult enough. To master two in a lifetime hardly seems possible, except that Reinhard Selten has done it.

References

1. Ball, Sheryl B. and Paula-Ann Cech. (1990), "The What, When, and Why of Picking a Subject Pool," Boston University working paper.
2. Banks, Jeff, Colin F. Camerer, and David Porter. (1994), "Experimental Tests of Refinements of Nash equilibrium in signaling games," *Games and Economic Behavior*, 6: 1-31.
3. Camerer, C. F., Eric J. Johnson, Talia Rymon, and Sankar Sen. (1994), "Cognition and Framing in Sequential Bargaining for Gains and Losses," in K. Binmore, A. Kirman, and P. Tani (Eds.), *Frontiers of Game Theory*, Cambridge: MIT Press, 27-47.
4. Camerer, Colin F., and Richard H. Thaler. (1995), "Anomalies: Ultimatums, dictators, and manners," *Journal of Economic Perspectives*, 9: 209-219.
5. Cox, James C., and R. Mark Isaac. (1986), "Experimental economics and experimental psychology: Ever the twain shall meet?" in A.J. MacFadyen and H. W. MacFadyen (Eds.), *Economic Psychology: Intersections in Theory and Application*, Amsterdam: North-Holland.
6. Davis, Doug L., and Charles Holt. (1992), *Experimental economics*. Princeton: Princeton University Press.
7. Deng, Gang, Robert Franciosi, Praveen Kujal, Roland Michelitsch, and Vernon Smith. (1994), "Fairness: Effects on temporary and equilibrium prices in posted offer markets," unpublished.
8. El-Gamal, Mahmoud, and Thomas R. Palfrey. "Economical experiments: Bayesian efficient experimental designs," *International Journal of Game Theory*, in press.
9. Friedman, Daniel and Shayam Sunder. (1994), *Experimental Methods: A Primer for Economists*. Cambridge: Cambridge University Press.
10. Guth, Werner, Rolf Schmittberger, and Bernd Schwarze. (1982), "An experimental analysis of ultimatum bargaining," *Journal of Economic Behavior and Organization*, 3: 367-388.
11. Hey, John D. (1991), *Experiments in Economics*. Cambridge, UK: Blackwell.
12. Hoffman, Elizabeth, Kevin McCabe, Keith Shachat, and Vernon L. Smith . (1994), "Preferences, property rights and anonymity in bargaining games," *Games & Economic Behavior*, 7: 346-380.
13. Kachelmeier, Steven, S. Limberg ,and M. Schadewald. (1991), "A laboratory market examination of the consumer price response to information about producers' costs and profits," *The Accounting Review*, 66: 694-717.
14. Kahneman, Daniel, Jack Knetsch, and Richard Thaler. (1986), "Fairness as a constraint on profit seeking: Entitlements in the market", *American Economic Review*, 76: 728-741.
15. Kagel, John H., Battalio, Ray C., and Leonard Green. (1995), *Economic Choice Theory: An Experimental Analysis of Animal Behavior*. Cambridge: Cambridge University Press.
16. Kagel, John L. and Alvin E. Roth (Eds.). (1995), *Handbook of Experimental Economics*,

Princeton: Princeton University Press.

17. Polzer, Jeff T., Margaret A. Neale, and Patrick O. Glenn. (1993), "The effects of relationships and justification in an interdependent allocation task," *Group Decision and Negotiation*, 2: 135-148.
18. Schotter, Andrew, Keith Weigelt, and Charles Wilson. (1994), "A laboratory investigation of multiperson rationality and presentation effects," *Games and Economic Behavior*, 6: 445-468.
19. Smith, Vernon L. (1976), "Experimental economics: Induced value theory," *American Economic Review Papers and Proceedings*, 66: 274-279.
20. Smith, Vernon L. (1982), "Microeconomic systems as an experimental science," *American Economic Review*, 72: 923-955.
21. Smith, Vernon L., and James M. Walker. (1993), "Monetary rewards and decision cost in experimental economics," *Economic Inquiry*, 31: 245-261.